Learning to detect a tone in unpredictable noise

Pete R. Jones[1], **David R. Moore**[1], **Daniel E. Shub**[2], **and Sygal Amitay**[1]

[1]*MRC Insitute of Hearing Research*
University Park, Nottingham, UK, NG7 2RD
p.r.jones@ucl.ac.uk, david.moore2@cchmc.org, sygal@ihr.mrc.ac.uk

[2]*School of Psychology, University of Nottingham*
University Park, Nottingham, UK, NG7 2RD
daniel.shub@nottingham.ac.uk

Abstract: Eight normal-hearing listeners practiced a tone-detection task in which a 1 kHz target was masked by a spectrally unpredictable multitone complex. Consistent learning was observed, with mean masking decreasing by 6.4 dB over five sessions (4500 trials). Reverse-correlation was used to estimate how listeners weighted each spectral region. Weight-vectors approximated the ideal more closely after practice, indicating that listeners were learning to attend selectively to the task relevant information. Once changes in weights were accounted for, no changes in internal noise (psychometric slope) were observed. We conclude that this task elicits robust learning, which can be understood primarily as improved selective-attention.

PACS numbers: 43.66.Ba, 43.66.Dc

1. Introduction

Detection thresholds for a fixed-frequency sinusoid deteriorate by 20–50 dB in the presence of spectrally unpredictable multitone complex (Kidd et al., 2007). Such masking cannot be explained by overlapping activity in peripheral auditory filters, since it occurs even when the masker is spectrally distal ('across-channel interference'), or energetically weak ('excess-additivity'). Instead, it is driven by higher order factors, including the degree of masker uncertainty (Neff and Callaghan, 1988), or the amount of target-masker similarity (Lee and Richards, 2011).

Whether listeners can learn to reduce such 'informational' masking is of general interest, since it is a prominent source of discomfort and dissatisfaction, particularly among hearing-impaired listeners. However, the extent of learning on such tasks has tended to be obscured by the use of highly experienced listeners. Two studies by Neff and colleagues *did* explicitly examine practice effects (Neff and Callaghan, 1988; Neff and Dethlefs, 1995), and these concluded that performance remained "remarkably stable" across sessions. However, in neither study were listeners naïve to the task. Thus, the listeners in Neff and Dethlefs (1995) completed 600 practice trials prior to testing, while those in Neff and Callaghan (1988) had extensive experience (> 10 h) of related masking tasks. Moreover, listeners generally completed fewer than 1000 trials. Given that auditory learning tends to be greatest early in training (Hawkey et al., 2004), and may extend over many thousands of trials, the full extent of learning remains uncertain.

The mechanisms subserving learning also remain unclear, and understanding them is important for the design of effective training schedules. One possibility is that levels of unpredictable masking are determined by the width of the listener's 'window of attention' (i.e., the spectral range over which auditory filter activity is integrated; Lutfi, 1993). Learning may therefore represent a narrowing of this window, with listeners learning to give weight only to the target region. Alternatively, learning may be the result of reduced internal noise (i.e., reduced variability in the listener's decision variable), as has been suggested previously (Jones et al., 2013). We evaluated both of these possibilities in the present study, using a two-step procedure developed by Berg (2004). Firstly, reverse-correlation was used to estimate the relative weight the listener gave to each spectral region. The observed weights were compared to the ideal to derive a measure of efficiency. Secondly, the estimated weights were used to derive the listener's trial-by-trial decision variable, DV, and a psychometric curve was fitted to the probability of responding 'Interval 2' as a function of DV. Since the goodness of the weighting strategy was partialled out, the slope parameter could be interpreted as an unambiguous index of internal noise.

2. Methods

2.1. Listeners

Eight listeners (five female) participated. They were aged 19–26, and had no prior experience of psychophysics. All had audiometrically normal hearing ($\leq$ 20 dB HL bilaterally at 0.25–8 kHz octaves). Participants were recruited through advertisements placed around the Nottingham University campus, and received £7.5/hour.

2.2. Task & Procedure

The task was two-alternative forced-choice [2AFC] tone detection, in which participants were asked to "pick the interval containing the target tone". Each trial consisted of two 300 ms observation intervals, separated by a 500 ms interstimulus interval. Responses were unspeeded, and were followed by 250 ms of visual feedback.

In each block, a two-down one-up adaptive track was used to derive an estimate of the listener's 70.7% detection limen (DL), either in noise or in quiet. The level of the target tone was initialized at 60 dB SPL. It was adapted in steps of 8 dB until the second reversal, and 2 dB thereafter. Each block consisted of 50 trials. The number of trials was fixed rather than the number of reversals in order to ensure that all listeners received the same amount of practice. Before each block listeners were presented with the target in quiet as a reminder.

Each session lasted approximately 45 minutes, and consisted of 16 noise blocks and two quiet blocks. The 18 blocks were presented in random order, with a rest break after the tenth block. Listeners completed five sessions over five consecutive days (4500 trials). Initially, participants also completed one practice trial in quiet and three practice trials in noise. To highlight the task demands, during this practice the stimuli durations were increased to 800 ms, and noises were attenuated to 50 dB SPL.

2.3. Stimuli & Apparatus

The target was always a 1 kHz sinusoid, randomly assigned to one of two observation intervals. In noise blocks, a 30-component multitone complex was also presented in each interval. All stimuli were 300 ms in duration, including 10 ms cos^2 on/off ramps, and were presented diotically over Sennheiser HD 25-I headphones.

The frequency, phase, and amplitude of the noise components were independently randomized on every presentation. Phases were randomly drawn from a rectangular distribution. Amplitudes were randomly drawn from a Rayleigh distribution, and normalized so that the total masker level was always 60 dB SPL. Frequencies were randomly drawn without replacement from 715 candidates, log-distributed between

223–4490 Hz, excluding a third-octave notch geometrically centered on the target frequency (891–1120 Hz). This notch served to minimize energetic masking, and was slightly greater than the average equivalent rectangular band [ERB] at 1 kHz (Glasberg and Moore, 1990).

Stimuli were digitally synthesised in Matlab v7.4 (2007a, The MathWorks, Natick, MA) using 44.1 kHz sampling and 24-bit quantization. Digital-to-analog conversion was performed by a PCI sound card (Darla Echo; Echo Digital Audio Corporation, Carpinteria, CA). Listeners responded via a button box, and were tested individually in a double-walled sound-attenuating booth.

2.4. Measures

DLs were calculated independently for each track as the mean target level (dB) at the last four reversals. With noise blocks, the amount of masking was computed by subtracting the mean DL in quiet (averaged over all blocks).

Relative weights were calculated via multiple logistic regression, as per Alexander and Lutfi (2004). The dependent variable was the listener's binary response ('Interval 1' or 'Interval 2'). The independent variables were the differences in (dB) level between the corresponding spectral region in each observation interval. In instances where there was no energy in a spectral region, the level was set to 0 dB. The weights, ω, were the regression coefficients, normalized so that their magnitudes summed to 1. As in Alexander and Lutfi (2004), the efficiency of the weight strategy was calculated as one minus the root mean square (RMS) difference between observed and ideal weights. The ideal strategy was to assign a weight of unity to the target bin and zero-weight elsewhere, which would yield an efficiency value of 1.

The magnitude of the internal noise, σ_{int}, was calculated as the standard deviation of a zero-mean cumulative normal distribution, fitted to the binned probability of a listener responding 'Interval 2' as a function of the estimated decision variable, DV. The DV was defined as $\sum_{i=1}^{n} \omega_i \Delta L_i$, where ΔL_i represents the energetic level difference (in dB) in the ith spectral bin, and ω_i is the corresponding relative weight coefficient. Psychometric fits were made using PSIGNIFIT v2.5.6: a Matlab toolbox which implements the maximum-likelihood method described by Wichmann and Hill (2001).

3. Results and Discussion

3.1. Learning

DLs in quiet and in noise are plotted for individuals in Fig 1. In quiet no learning was observed, with all individuals well-described by linear regressions with near-zero slope [$\beta_\mu = 0.01$; all $p > 0.05$]. In contrast, substantial learning was observed in the masked condition. Linear fits yielded significant negative slopes for all but two (L5, L7) listeners [$p < 0.05$], with improvement rates ranging from -0.02 to -0.18 dB/block. Some data were better fit by broken-stick functions, suggesting a short initial phase of rapid learning followed by a protracted period of more gradual learning. However, the improvements in fit were small [$\Delta_{r^2}/\Delta_{d.f.} < 1$] for all but L5, indicating that learning may be more gradual than in more basic auditory tasks, such as frequency discrimination (Hawkey et al., 2004). Grand mean masking decreased from 44.3 dB in session one to 38.0 dB in session five [$t(7) = 4.09, p = .005$].

We conclude that the ability to detect a tone in unpredictable noise improves robustly with practice. Furthermore since the performance of some listeners (e.g., L1 and L4) had not plateaued by the end of the study, further learning may have been possible. This release from across-channel interference is consistent with Buss (2008), who found that six of eight listeners improved with practice on an analogous intensity-discrimination in unpredictable-noise task. However, widespread learning is contrary to Neff and Dethlefs (1995), where tone-in-unpredictable-noise DLs appeared 'remarkably stable' in most listeners. This difference is likely due to the listeners in Neff and Dethlefs (1995) having completed 600 trials practice prior to training. Accordingly, when the first 600 trials were excluded in the present study, regression slopes were significantly shallower [$t(7) = -2.43$, $p = 0.046$], and failed to differ from zero (no-learning) in 5 of 8 listeners [five $p > 0.05$].

Consistent with previous reports, there was substantial individual variability in masking: 36.6–58.0 dB in session one, 31.6–50.0 dB in session five. Previous authors have wondered whether such individual differences can be reduced by training (e.g., Durlach et al., 2003). As in Neff and Callaghan (1988), we found little evidence of that here; between-subject variation in masking was approximately constant across all five sessions, with the greatest variability occurring in session four where group mean masking was actually lowest. There was some indication that variability within listeners (intra-session masking S.D.) decreased with practice, but this change was not significant [$t(7) = 2.27; p = .058, n.s.$].

3.2. Mechanisms of learning

Estimates of weights are shown for individuals in Fig 2. Most listeners gave greatest weight to the target bin (1 kHz), but also appeared to weight irrelevant spectral information. As in some previous reports, this overweighting was particularly prevalent at the upper fringe of the stimulus (c.f., Watson et al., 1976), suggesting that a selective-attenuation of these high frequency noise components alone may provide a substantive release from masking. A repeated measures ANOVA yielded a significant main effect of session on weight efficiency [$F(4, 28) = 3.48$, $p = 0.020$], indicating that listeners' weighting strategies improved with practice. Thus, after practice listeners predicated their decisions more selectively on the target spectral region.

To evaluate changes in internal noise, cumulative Gaussians were fitted to listeners' responses as a function of the estimated trial-by-trial DV. Group-mean estimates of σ_{int} did not systematically vary across session [$F(4, 28) = 0.20, p = 0.937, n.s.$], indicating that internal noise magnitude was not diminished by practice. However, as shown in Fig 3, there was variability between listeners. For example, listeners L1 and L4 exhibited a marked decrease in internal noise, as indicated by steeper psychometric slopes in session five. Conversely, listener L8 shows very little change (but did exhibit substantial learning; c.f., Fig 1).

Alternate psychometric fits were also made assuming ideal weights throughout (i.e., P('Interval 2') as a function of target level). These models did indicate a reduction in internal noise [$F(4, 28) = 5.49, p = 0.002, \eta_p^2 = 0.44$]. However they gave a markedly poorer account of the raw data, with significantly greater deviance between model and data [rm-ANOVA; $F(1, 63) = 11.02, p = 0.013, \eta_p^2 = 0.61$]. Thus, we favor the combined weight+internal noise account here, both as it provides empirically stronger fits, and because of its potential to predict how performance will change depending on how spectral composition of the stimulus varies.

In summary, training produced significant improvements in weight efficiency but not internal noise – listeners learned with practice to attend selectively to the task-relevant information. Finally, to examine whether these changes differed in *effect* as well as significance, Monte Carlo simulations were run using the observed group-mean changes in weight efficiency and internal noise. These simulations followed the same test procedure as human listeners (i.e., same *n* trials, *n* listeners, etc.), and the simulations similarly responded to the greatest weighted-sum-energy. Thus, the decision rule was to respond 'Interval 1' only when $\left[\sum_{i=1}^{n} \omega_i \Delta L_i + \sigma_{int}/\sqrt{2}\right] > 0$.

When internal noise was held constant at its mean session 1 value, and weight-efficiency was varied from its session 1 to session 5 value, masking decreased by an amount similar to the listeners [$Sim : -7.2$ dB; $Listener : -6.7$ dB]. Conversely, varying internal noise and holding weight-efficiency constant produced no significant change in masking [$p = .399$]. These results suggest that improvements in selective attention primarily determine learning on this task.

4. Conclusions

(1) Masking by unpredictable noise is reduced by practice in most listeners. Some of this learning occurs rapidly, within the first 600 trials, but learning may continue for several thousand trials.

(2) Improvements in weight efficiency underlie learning on this task, with listeners becoming better-able to attend selectively to the task-relevant information.

Acknowledgments

This work was supported by the Medical Research Council, UK (Grant: U135097130).

References and links

Alexander, J. M. and Lutfi, R. A. (**2004**). Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights. *J. Acoust. Soc. Am.*, 116:2234–2247.

Berg, B. G. (**2004**). A molecular description of profile analysis: Decision weights and internal noise. *J. Acoust. Soc. Am.*, 115(2):822–829.

Buss, E. (**2008**). Across-channel interference in intensity discrimination: The role of practice and listening strategy. *J. Acoust. Soc. Am.*, 123(1):265–272.

Durlach, N. I., Mason, C. R., Kidd Jr, G., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). Note on informational masking (l). *J. Acoust. Soc. Am.*, 113(6):2984–2987.

Glasberg, B. R. and Moore, B. C. J. (**1990**). Derivation of auditory filter shapes from notched-noise data. *Hear. Res.*, 47(1-2):103–138.

Hawkey, D. J. C., Amitay, S., and Moore, D. R. (**2004**). Early and rapid perceptual learning. *Nat. Neurosci.*, 7(10):1055–1056.

Jones, P. R., Shub, D. E., Moore, D. R., and Amitay, S. (**2013**). Reduction of internal noise in auditory perceptual learning. *J. Acoust. Soc. Am.*, 133(2):970–981.

Kidd, Jr., G., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (**2007**). Informational masking. In Christophe Micheyl, S. A. S. and Oxenham, A. J., editors, *Auditory perception of sound sources*, pages 143–189. Springer Science+Business Media, New York, New York.

Lee, T. Y. and Richards, V. M. (**2011**). Evaluation of similarity effects in informational masking. *J. Acoust. Soc. Am.*, 129(6):EL280–EL285.

Lutfi, R. A. (**1993**). A model of auditory pattern analysis based on component-relative-entropy. *J. Acoust. Soc. Am.*, 94(2):748–758.

Neff, D. L. and Callaghan, B. P. (**1988**). Effective properties of multicomponent simultaneous maskers under conditions of uncertainty. *J. Acoust. Soc. Am.*, 83(5):1833–1838.

Neff, D. L. and Dethlefs, T. M. (**1995**). Individual differences in simultaneous masking with random-frequency, multicomponent maskers. *J. Acoust. Soc. Am.*, 98(1):125–134.

Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). Factors in the discrimination of tonal patterns. ii. selective attention and learning under various levels of stimulus uncertainty. *J. Acoust. Soc. Am.*, 60(5):1176–1186.

Wichmann, F. A. and Hill, N. J. (**2001**). The psychometric function: I. fitting, sampling, and goodness of fit. *Atten. Percept. Psychophys.*, 63(8):1293–1313.

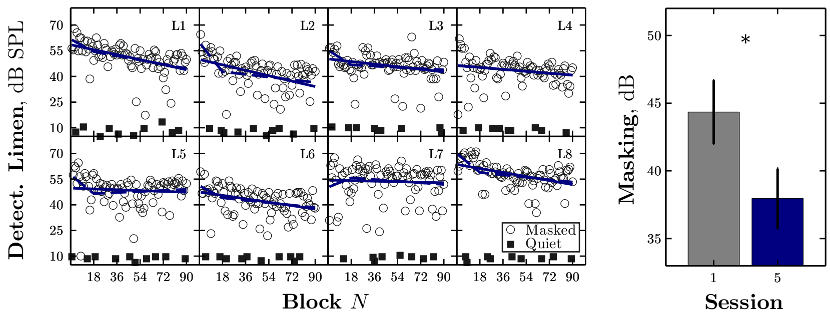

Fig. 1. (Left) Detection limens for individuals, as a function of block. DLs in quiet and in noise are shown by filled squares and open circles, respectively. Solid lines represent least square linear fits to the noise data. Dashed-lines denote broken-stick fits, inflected after session 1. (Right) Group-mean (± 1 SE) change in masking. (Color online).

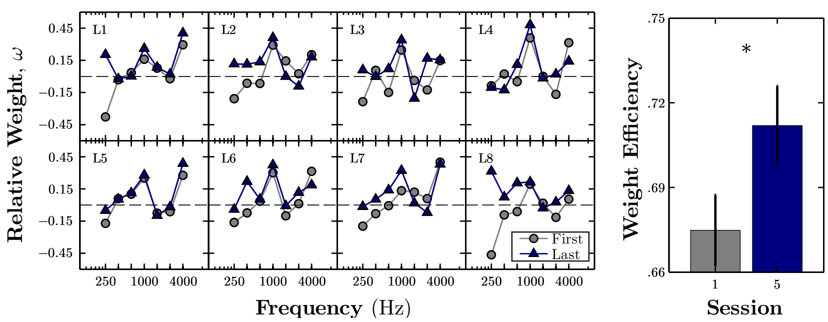

Fig. 2. (Left) Individual weights, for the first (grey circles) and last (blue triangles) session. Each point represents the geometric center of a third-octave spectral bin. The target signal was always a 1 kHz sinusoid, so the optimal strategy was to always give unit weight to the 1 kHz bin, and zero weight elsewhere. (Right) Group-mean (± 1 SE) change in weight efficiency. (Color online).

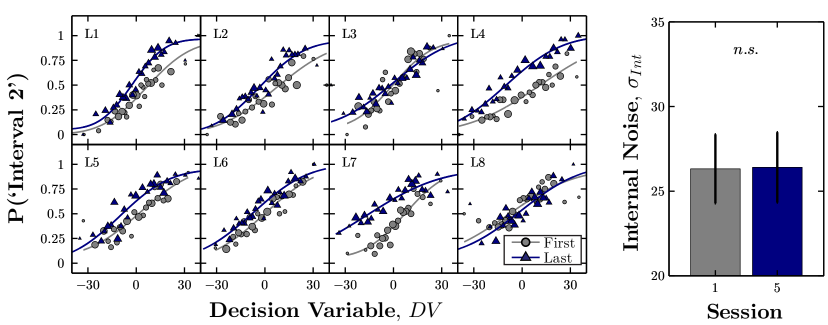

Fig. 3. (Left) Individual psychometric fits, for the first (grey circles) and last (blue triangles) session, relating the probability of responding 'Interval 2' to the estimated trial-by-trial decision variable. The standard deviation parameter, σ, of the best cumulative Gaussian fit was taken as a measure of internal noise magnitude. (Right) Group-mean (± 1 SE) change in internal noise magnitude. (Color online).